Эта программа, точнее, те ее части, которые относятся к более конкретизированным проблемам, разрабатывается коллективом математиков и экономистов в рамках проекта, получившего грант РФФИ – Урал. Разработки основаны на имеющемся у нас богатом опыте применения математических средств к решению указанных проблем.

Мы говорили об экономическом и природном потенциале территорий с точки зрения первичного, грубого упорядочения, структуризации траекторий. Это основа для обнаружения закономерностей, которые могут быть использованы в управлении территориями. Но существуют ли закономерности объективно или мы их навязываем, набрасывая сетку понятий на хаотическую действительность? По нашему мнению, при ответе на этот вопрос следует избегать крайностей – нужна некоторая средняя точка зрения.

Абсолютный хаос нельзя наделить каким-либо смыслом. Но, к счастью, абсолютный хаос так же нереален, как и абсолютная детерминированность. И последняя тоже не имеет смысла, т. к. полностью исключает выбор и свободу творчества. Надо идти по узкой меже, разделяющей эти области. Это относится и к математическим моделям: они не могут быть бесформенными, но они в то же время должны быть «мягкими», нестационарными, открытыми. Они должны отталкиваться от предварительно упорядоченного материала наблюдений. Найденные в рамках этого материала эмпирические закономерности позволяют обнаруживать скрытые факторы, влияющие на социальную и экономическую динамику. Это предполагает многомерный взгляд на мир, который на самом деле является (актуально или потенциально) бесконечномерным. Впрочем, вряд ли кто-нибудь уверенно может сказать, каким на самом деле является мир.

П. И. Браславский

МОРФОЛОГИЧЕСКИЙ СТРОЙ ФУНКЦИОНАЛЬНЫХ СТИЛЕЙ

(на материале документов Internet)

Введение

Данная статья содержит результаты, полученные в рамках разработки процедуры автоматической классификации текстов по стилям. Стилистическая классификация, в свою очередь, рассматривается как одно из средств повышения эффективности поиска информации в Internet [2–4], при этом морфологические характеристики в процедуре классификации имеют ключевое значение.

Дополнительным стимулом в данной работе было желание продемонстрировать возможность использования наполнения Internet

в лингвистических исследованиях. Обращаясь к сети, исследователь получает доступ к неограниченному объему самых разнообразных текстов в электронном виде. Так, например, масштабные исследования разговорной речи всегда сдерживались отсутствием достаточного количества опытного материала в форме, удобной для обработки. Сегодня чаты, гостевые книги, форумы, а также архивы личной переписки по электронной почте и общения по ICQ могут предоставить такой материал в избытке.

За основу мы взяли функционально-стилевую концепцию, которая хорошо разработана и обоснована в отечественном языкознании [11, 12, 14]. Исходным положением концепции является зависимость стиля речи от выполняемой им коммуникативно-общественной функции, от задач общения в соответствующей сфере. Обычно различают пять функциональных стилей речи (в порядке убывания «нормативности»): *официально-деловой, научный, публицистический, художественный, разговорный* (исходя из прикладного характера задачи, мы рассматриваем *художественный стиль* наряду с другими, не учитывая его особый статус в системе функциональных стилей).

Исследования функциональных стилей с использованием статистических методов проводились начиная с 60-х годов. В работах [1, 5–10, 12–17] можно найти количественные характеристики морфологии стилей речи разной степени детализации. Недостаток большинства этих источников – использование для анализа выборок небольшого объема (часто трех – пяти текстов). Кроме того, не всегда понятно, какие именно тексты послужили материалом для исследования и какая методика использовалась. Практически нигде не удается найти интегральную картину распределения классов слов по стилям: обычно одновременно рассматривается не более трех стилей. «Частотный словарь» [17] лишен этих минусов (общий объем обработанного материала – 1 056 382 слова), однако деление на стили (жанры) представляется не очень логичным: *художественная проза, драматургия, газетно-журнальные и научно-публицистические тексты*.

Речь является динамической системой, и значительные стилевые изменения могут происходить на относительно коротких временных промежутках (см. работы [7, 8, 13], посвященные исследованию динамики функциональных стилей). Определение стилистических особенностей «сетевых» текстов интересно еще и потому, что сегодня бумагу и ручку (печатную машинку) заменяет компьютер, а Internet – фактор не менее значительный, чем печатный станок пятьсот лет назад. Смена способа материальной фиксации текстов безусловно влияет на их стиль (можно вспомнить происхождение самого слова «стиль» – от лат. *stilus*, *stylus* – остроконечная палочка для письма).

Опытный массив текстов

Взятая за основу функционально-стилевая концепция определила наш подход к формированию массива текстов для анализа. В опытном массиве каждый *стиль* представлен наиболее типичным *жанром*; задача представления жанрового разнообразия в пределах функционального стиля не ставилась. Очевидно, что составить репрезентативную коллекцию всего стиля, которая учитывала бы количественные соотношения между различными жанрами, их вклад в «общую картину» стиля, весьма затруднительно.

Такой переход (от стиля к жанру) вполне отвечает прикладным целям нашего исследования.

Официально-деловой стиль представлен в опытном массиве текстами 50 законов Российской Федерации. Эти документы были отобраны из юридической базы данных «Консультант Плюс» (www.consultant.ru). Дополнительным критерием отбора была длина текста. Например, были отсеяны законы о ратификации договоров, которые обычно содержат 2–3 строчки.

В коллекцию текстов *научного стиля* вошли 54 статьи по физике, математике, химии, биологии и инженерным наукам. Практически все электронные версии научных статей, размещенные в Internet, имеют печатные аналоги.

Публицистический стиль, напротив, представлен только Internet-журналистикой. В качестве представителей этого стиля мы взяли статьи на общественно-политические темы, опубликованные в период с декабря 1999 по февраль 2000 года на трех новостных веб-сайтах: Gazeta.ru (27), Vesti.ru (28) и Polit.ru (6) – всего 61 статья.

Художественный стиль в нашем исследовании представлен 79 рассказами участников конкурса сетевой литературы «Тенета-98» (www.teneta.rinet.ru/1998/rasskaz/). Нам представляется закономерным использовать для анализа произведения, которые увидели свет в Internet, а не литературную классику.

Основной объем текстов *разговорного стиля* принадлежит екатеринбургскому чату «На Плотинке» (www1.ekaterinburg.com/leisure/chat/) – 42 фрагмента, каждый из которых содержит ровно 100 сообщений. Кроме того, два фрагмента взято с чата «Сайберия» (www.son.ru/chat/) и четыре – с чата «В пещере у монстра» (cave.extrim.ru). В данном случае объем каждого фрагмента – примерно 2–3 экрана. Дополнили коллекцию образцов разговорного стиля 13 листингов диалогов (14 разных участников), которые велись с помощью программы ICQ («аська»). Таким образом, всего был использован 61 фрагмент.

Все функциональные стили рассматриваются изолированно, поэтому некоторые различия в объемах текстов каждого стиля несущественны. При этом массив текстов достаточно представителен (как по отдельным стилям, так и в целом), чтобы вычисленные параметры были значимы. Общий объем массива – 305 текстов.

Методика обработки

Анализу подвергались текстовые документы (plain text) и документы HTML в Windows-кодировке. Документы Word и Adobe Acrobat (PDF) предварительно конвертировались в текстовые файлы.

Для автоматического определения грамматических характеристик слов использовался модуль морфологического анализа LINGUIST компании «Агама» (www.agama.com). По информации разработчиков основной словарь модуля морфологического анализа и синтеза позволяет распознавать более четырех миллионов словоформ. Модуль выполнен в виде динамической библиотеки Windows.

По аналогии с предыдущими исследованиями морфологии функциональных стилей и в соответствии с возможностями модуля LINGUIST в качестве самостоятельных морфологических классов были выделены:

1) существительные,
2) прилагательные,
3) местоимения,
4) числительные,
5) наречия,
6) глаголы,
7) причастия,
8) деепричастия,
9) предлоги,
10) союзы,
11) частицы,
12) междометия,
13) прочие.

К существительным мы также относили следующие категории модуля морфологического анализа: *имена собственные, отчества, фамилии, географические названия, аббревиатуры*. В разряд «Прочие» попали слова, которые модуль морфологического анализа отнес к *предикативам* или *вводным словам*.

Причастия и деепричастия выделены в самостоятельные классы, в силу их стилистической окрашенности. Краткие прилагательные, хотя и имеют выраженную стилистическую окраску, не выделены в самостоятельную группу, т. к. не учитываются модулем морфологического анализа.

В каждом тексте анализировались первые 1000 русских слов, а также слова до конца текущего предложения, или весь текст, если его длина меньше 1000 слов.

Словом считалась последовательность русских букв (которая может содержать внутри себя дефис) между двумя разделителями. Слова, содержащие цифры или латинские буквы, не анализировались. Словоформы, для которых модуль не возвращает ни одной нормальной формы, не учитывались.

Итог обработки отдельного текстового фрагмента – строка значений, каждое из которых соответствует доле части речи в тексте.

Ясно, что данные, полученные таким образом, не являются абсолютно точными. Сделав ставку на автоматическую обработку материала большого объема, приходится отказаться от учета грамматической омонимии. Поэтому, например, слова *стекло, падали* всегда относятся к существительным, как и *рабочий, учащийся*; а *печь, течь* – к глаголам. Кроме того, из-за переносов или вставки невидимых символов и тегов HTML в слово могут возникать ошибки определения границ слова.

Результаты

В соответствии с методикой подсчета параметров было обработано 305 фрагментов текста. Общий объем обработанного материала – 239 696 слов, по 227 257 из них модулем морфологического анализа были построены нормальные формы (установлены грамматические характеристики). Доля слов из русских букв, для которых модулем морфологического анализа не было построено ни одной нормальной формы, составляет 6,1 % (со значительным разбросом по отдельным стилям: разговорный – 15 %, художественный – 2,7 %, публицистический – 4,6 %, научный – 6,9 %, официально-деловой – 1,9 %).

Результаты обработки данных по каждому стилю и части речи приведены в табл. 1: среднее ($x_{ср}$), минимальное (min) и максимальное (max) значения, стандартное отклонение (S).

В целом полученные данные неплохо согласуются с результатами предыдущих исследований. Результат одновременного рассмотрения пяти стилей – монотонный рост средних долей существительных и прилагательных и монотонное же уменьшение долей местоимений, наречий, глаголов

Таблица 1

Статистика частей речи по стилям

Параметры	Существительные	Прилагательные	Местоимения	Числительные	Наречия	Глаголы	Причастия	Деепричастия	Предлоги	Союзы	Частицы	Междометия	Прочие
Разговорный стиль (61 фрагмент, 30601 слово)													
min	0,074	0,000	0,105	0,000	0,039	0,109	0,005	0,000	0,024	0,021	0,132	0,000	0,000
max	0,298	0,000	0,222	0,009	0,102	0,219	0,050	0,042	0,084	0,085	0,276	0,045	0,043
$x_{ср}$	**0,194**	**0,000**	**0,161**	**0,002**	**0,068**	**0,167**	**0,028**	**0,006**	**0,051**	**0,050**	**0,210**	**0,016**	**0,013**
S	0,040	0,000	0,027	0,002	0,017	0,024	0,011	0,007	0,013	0,013	0,031	0,009	0,008
Художественный стиль (79 рассказов, 73739 слов)													
min	0,140	0,022	0,059	0,000	0,023	0,091	0,016	0,001	0,031	0,014	0,068	0,000	0,000
max	0,351	0,106	0,227	0,019	0,118	0,239	0,074	0,028	0,085	0,063	0,242	0,018	0,014
$x_{ср}$	**0,243**	**0,063**	**0,126**	**0,006**	**0,065**	**0,162**	**0,045**	**0,013**	**0,055**	**0,037**	**0,158**	**0,003**	**0,006**
S	0,049	0,020	0,039	0,004	0,017	0,027	0,013	0,007	0,010	0,011	0,030	0,003	0,003
Публицистический стиль (61 статья, 34939 слов)													
min	0,265	0,061	0,036	0,000	0,022	0,079	0,030	0,000	0,023	0,023	0,068	0,000	0,000
max	0,410	0,175	0,119	0,027	0,077	0,165	0,102	0,020	0,084	0,061	0,221	0,007	0,020
$x_{ср}$	**0,335**	**0,107**	**0,075**	**0,007**	**0,049**	**0,120**	**0,066**	**0,009**	**0,058**	**0,038**	**0,130**	**0,000**	**0,007**
S	0,034	0,024	0,019	0,005	0,012	0,019	0,017	0,005	0,013	0,008	0,029	0,001	0,004
Научный стиль (54 статьи, 47264 слова)													
min	0,245	0,074	0,019	0,000	0,004	0,047	0,044	0,002	0,023	0,010	0,054	0,000	0,000
max	0,499	0,199	0,081	0,019	0,087	0,129	0,141	0,041	0,103	0,117	0,156	0,023	0,029
$x_{ср}$	**0,396**	**0,130**	**0,047**	**0,005**	**0,029**	**0,090**	**0,091**	**0,017**	**0,061**	**0,033**	**0,090**	**0,001**	**0,008**
S	0,054	0,028	0,013	0,004	0,016	0,020	0,021	0,010	0,015	0,022	0,022	0,004	0,006
Официально-деловой стиль (50 законов, 33134 слова)													
min	0,427	0,075	0,010	0,000	0,000	0,015	0,046	0,000	0,010	0,000	0,024	0,000	0,000
max	0,608	0,278	0,052	0,061	0,037	0,100	0,140	0,026	0,087	0,032	0,136	0,000	0,018
$x_{ср}$	**0,497**	**0,184**	**0,029**	**0,009**	**0,008**	**0,048**	**0,091**	**0,005**	**0,046**	**0,009**	**0,071**	**0,000**	**0,002**
S	0,037	0,048	0,011	0,012	0,007	0,018	0,023	0,005	0,020	0,008	0,019	0,000	0,004

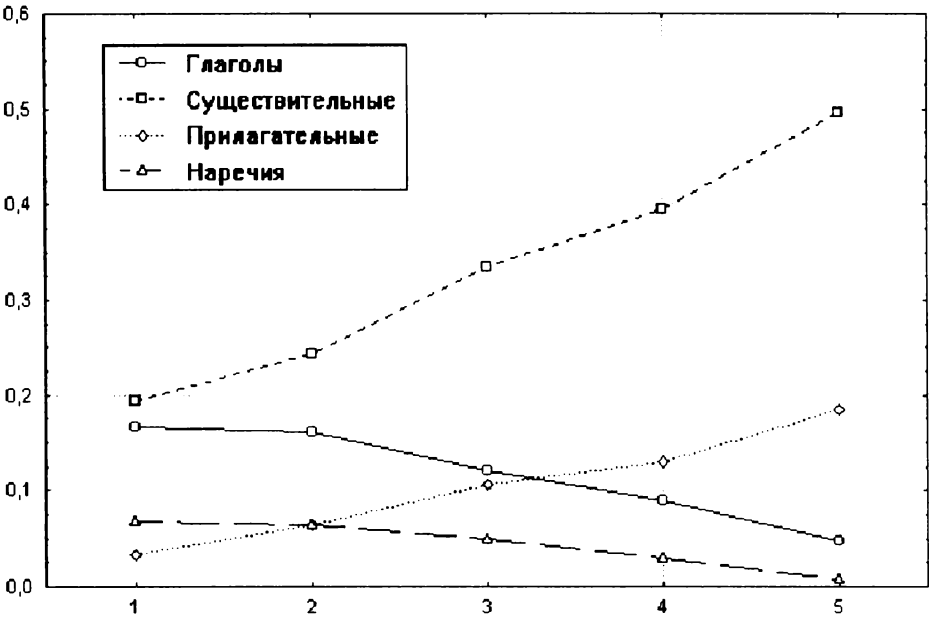
Таблица 2

Матрица корреляции

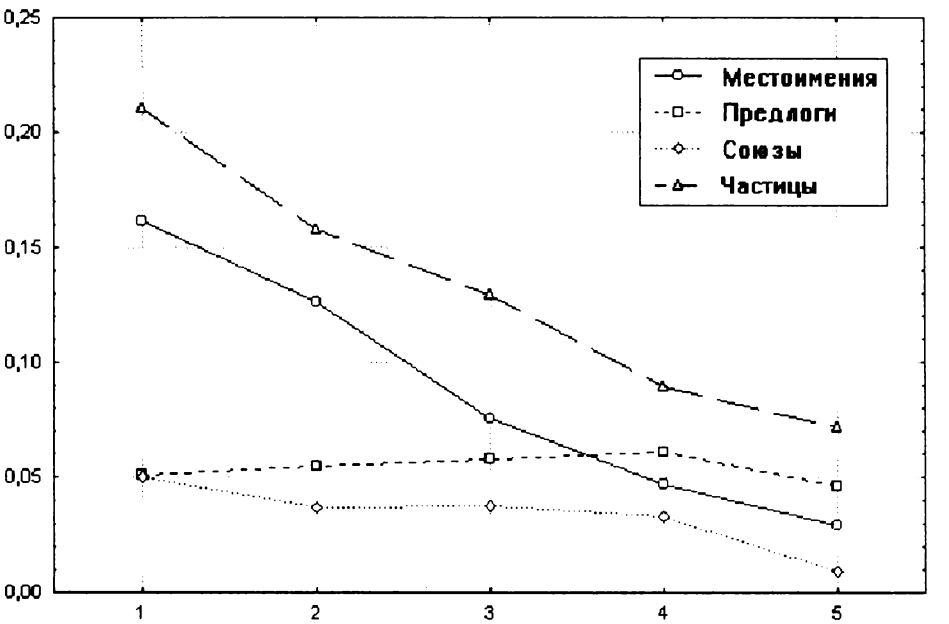
№ п/п	Части речи	1	2	3	4	5	6	7	8	9	10	11	12
1	Существительные	1,00	**0,85**	**-0,87**	0,21	**-0,85**	**-0,88**	0,77	-0,03	0,00	**-0,72**	**-0,86**	-0,54
2	Прилагательные	**0,85**	1,00	**-0,81**	0,09	**-0,75**	**-0,85**	0,67	-0,01	-0,11	-0,67	**-0,79**	-0,54
3	Местоимения	**-0,87**	**-0,81**	1,00	-0,21	**0,70**	**0,79**	**-0,78**	-0,08	-0,08	0,57	**0,77**	0,53
4	Числительные	0,21	0,09	-0,21	1,00	-0,15	-0,18	0,21	-0,07	0,09	-0,11	-0,20	-0,25
5	Наречия	**-0,85**	**-0,75**	**0,70**	-0,15	1,00	**0,80**	-0,69	0,08	0,02	0,63	**0,76**	0,38
6	Глаголы	**-0,88**	**-0,85**	0,79	-0,18	**0,80**	1,00	**-0,75**	0,03	0,02	0,62	**0,75**	0,46
7	Причастия	**0,77**	0,67	**-0,78**	0,21	-0,69	**-0,75**	1,00	0,04	0,13	-0,50	**-0,77**	-0,50
8	Деепричастия	-0,03	-0,01	-0,08	-0,07	0,08	0,03	0,04	1,00	0,09	0,14	-0,12	-0,23
9	Предлоги	0,00	-0,11	-0,08	0,09	0,02	0,02	0,13	0,09	1,00	0,02	-0,13	-0,06
10	Союзы	**-0,72**	-0,67	0,57	-0,11	0,63	0,62	-0,50	0,14	0,02	1,00	0,63	0,39
11	Частицы	**-0,86**	**-0,79**	**0,77**	-0,20	**0,76**	**0,75**	**-0,77**	-0,12	-0,13	0,63	1,00	0,61
12	Междометия	-0,54	-0,54	0,53	-0,25	0,38	0,46	-0,50	-0,23	-0,06	0,39	0,61	1,00

Примечание. Полужирным шрифтом выделены коэффициенты корреляции, по модулю большие или равные 0,70.

а



б



Средние доли частей речи по стилям:
1 – разговорный; 2 – художественный; 3 – публицистический;
4 – научный; 5 – официально-деловой

и частиц от разговорного к официально-деловому стилю – наглядно представлен на рисунке. При этом доля служебных частей речи (предлоги, союзы) мало варьируются от стиля к стилю (см. рис., б).

Анализ матрицы корреляции (табл. 2), вычисленной по всему корпусу текстов, позволяет выделить группу взаимосвязанных морфологических параметров: существительные, прилагательные, причастия, глаголы, наречия, местоимения, частицы. Это вполне объяснимый результат: каждое употребление существительного – это «повод» определить его прилагательным; действие, выраженное глаголом, можно охарактеризовать наречием; функция местоимения – заменять именные части речи и т. д. Матрица корреляции демонстрирует, что частота употребления слов одной части речи из этой группы во многом определяет частоту употребления остальных. Зная, например, долю существительных в тексте, мы можем делать достаточно точные прогнозы относительно доли прилагательных и глаголов. Это справедливо даже для текстов, сильно отличающихся по стилю.

Заключение

В качестве основных результатов исследования можно выделить следующие:

1. Морфологические параметры (частеречный спектр текста) принадлежат к важнейшим маркерам функционального стиля и поэтому могут быть использованы для автоматической классификации текстов по стилям.

2. Internet содержит текстовый материал всех функциональных стилей русской речи, что открывает новые перспективы для исследований общего и стилистико-сопоставительного характера. Полученные результаты говорят об эффективности метода, основанного на автоматической обработке больших объемов текстов.

3. Получены количественные характеристики морфологии частей речи функциональных стилей русской речи и проведена их статистическая обработка.

В заключение хочется выразить надежду, что представленные в статье результаты получат более содержательную интерпретацию с позиций стилистики.

* * *

Мы благодарим компанию «Агама» (www.agama.com) за предоставленный модуль морфологического анализа, а также Михаила Щекотилова за программную реализацию метода.

Литература

1. *Андреев Н. Д.* Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967.

2. *Браславский П. И.* Автоматическая классификация документов Internet по стилям: реализация макета [Электрон. ресурс]// Доклады V Рабочего совещания по электронным публикациям – EL-PUB-2000 / Новосибирск: ИВТ СО РАН. – Электрон. дан. – http://www.ict.nsc.ru/ws/el-pub-2000/29/. – 21.06.2000.

3. *Браславский* П. И. Использование стилистических параметров документа при поиске информации в Internet [Электрон. ресурс]// Доклады VI Рабочего совещания по электронным публикациям – EL-PUB-2001 / Новосибирск: ИВТ СО РАН. – Электрон. дан. – http://www.ict.nsc.ru/ws/elpub2001/1812/. – 25.04.2001.

4. *Браславский П. И.* Распознавание стилей речи применительно к информационному поиску: постановка задачи // Математические структуры и моделирование: Сб. науч. тр. Вып. 3 / Под ред. А. К. Гуца. Омск: Омский гос. ун-т, 1999. С. 134–140.

5. *Васильева А. Н.* Курс лекций по стилистике русского языка. Научный стиль речи. М., 1976.

6. *Головин Б. Н.* Язык и статистика. М., 1970.

7. Изменения в языке научной прозы/ О. Б. Сиротинина, С. А. Бах, В. А. Богданова и др. // Вопр. стилистики. Вып. 3. Саратов: Изд-во Сарат. ун-та, 1969. С. 37–55.

8. Изменения в языке публицистики (на материале международных обзоров) / О. Б. Сиротинина, С. А. Бах, В. А. Богданова и др. // Там же. С. 5–36.

9. *Кауфман С. И.* Из курса лекций по статистической стилистике. М., 1970.

10. *Клочкова Э. А.* О влиянии формы разговорной речи на распределение классов слов // Русская разговорная речь: Сб. науч. тр. Саратов: Изд-во СГУ, 1970. С. 126–134.

11. *Кожина М. Н.* К основаниям функциональной стилистики. Пермь, 1968.

12. *Кожина М. Н.* О речевой системности научного стиля сравнительно с некоторыми другими. Пермь, 1972.

13. Очерки истории научного стиля русского литературного языка XVIII–XX вв. / Под ред. М. Н. Кожиной: В 3 т. Т.1. Развитие научного стиля в аспекте функционирования языковых единиц различных уровней. Ч.1. Пермь, 1994.

14. Разговорная речь в системе функциональных стилей современного русского языка. Лексика / Под ред. О. Б. Сиротининой. Саратов: Изд-во Сарат. ун-та, 1983.

15. Русская разговорная речь. Фонетика. Морфология. Лексика. Жест. М., 1983.

16. *Сиротинина О. Б.* Современная разговорная речь и ее особенности. М., 1974.

17. Частотный словарь русского языка / Под ред. Л. Н. Засориной. М., 1977.

О. М. Ушакова

МИФ О ПРОКНЕ И ФИЛОМЕЛЕ В ПОЭЗИИ Т. С. ЭЛИОТА

Феномен возрождения общекультурного интереса к мифу на рубеже XIX–XX веков основательно исследован в отечественном и зарубежном литературоведении. История культуры XX века, современная культурная ситуация свидетельствуют о том, что процесс «ремифологизации» не завершен. Использование мифа как интегрального культурного пространства продолжается в новых формах и в постмодернистский период. Мифология, в силу своей универсальной природы, является по сей день адекватным языком описания вечных моделей личного и социального поведения, сущностных законов бытия, обладая при этом энергией, высвобождающей индивидуальные креативные импульсы. «Миф есть в словах данная чудесная личностная история»[1].

В литературе высокого модернизма была создана эстетически перспективная мифологическая поэтика, элементы которой до сих пор являются творчески продуктивными. Классические модернистские тексты «Улисс» Д. Джойса и «Бесплодная земля» Т. С. Эли-